# Final Project –

Due Wednesday, June 14th 2023 at 11:59 pm

# Introduction

During ENGR 103 you have learned programming skills applied to real-world energy efficiency and sustainability applications. It's now your turn to apply your skills to something you are interested in! And we are excited to help you explore.

You will find and analyze a dataset freely available online that is of interest to you using the programming tools you have learned in ENGR 103. Your project deliverables should include:

1. A ~2 pg .pdf file that introduces the dataset you chose and presents the synthesis of your analysis of the dataset in an easy and understandable way. This should be something you can explain to your grandparents. This should include your findings and figures that are labeled appropriately. For example, when the United Nations presents their data analysis on clean water and sanitation, they show graphics and statements that represent their analysis and synthesis of data in an understandable way – they are not showing us any code on their website. https://sdgs.un.org/goals/goal6
2. Your code:
    a. A .ipynb file that contains your main code in a well documented format (use comments).
    b. A .py file that contains your library of functions that you call in your main script that is also well documented.
    c. Your code should include elements described in the required list below.

# Required elements of your program

- Apply at least one *for* or *while* loop
- Apply at least one *if/elif/else* logical operators statements
- Sort the data in some way (biggest to smallest / by month / by year / by type / etc.)
- Perform at least three independent computations with the data (* / + - ** % //, etc.)
- Print at least 2 statements of output from your program to the terminal
- Create at least 3 plots showing your data and the analysis of your data that are well labeled and use different types of plotting styles
- Use the numpy, pandas, and matplotlib packages within your program
- Create a library that contains functions that you call in your main script for your analysis – you must have independent main and library scripts
- Use other coding elements that are applicable to your analysis (slicing, user input, 2D arrays, etc.

# How to choose a Dataset

**Ask yourself the following questions:**

1. Is the dataset relevant to your interests? What makes it interesting to you?
2. Is the dataset societally relevant? If not, could you pick something that is both interesting to you and that is societally relevant? e.g., energy, environment, health, social justice, resilience, safety, or if something else – please justify
3. Does the dataset come in either csv or txt format? Something easy for you to load into your IDE. Does the dataset consist of mostly numerical values and or short characters?

**If you don't know where to start… consider the following resources:**

World Health Organization Data: https://www.who.int/data (scroll down to the bottom to see what datasets they have)

NASA Global Climate Change Data: https://climate.nasa.gov/vital-signs/carbon-dioxide/ (the tabs, carbon dioxide, global temperature, arctic sea ice extent, ice sheets, sea level, and ocean heat content have datasets you can download and work with from NASA)

Ember Electricity Data: https://ember-climate.org/data/ (explore electricity data from around the world!)

Social Justice Data Repository: https://www.gvsu.edu/bigdata/social-justice-big-data-repository-29.htm

Department of Energy Datasets: https://www.energy.gov/data/open-energy-data

Kaggle Open Datasets: https://www.kaggle.com/datasets?sort=usability&fileType=csv&page=2 (this link sorts the data by csv file type and by usability – usability of 10 is the best!, you can continue your search on certain topics within the search bar, like 'environment')

Other Open Datasets: https://careers.uw.edu/blog/2021/10/05/21-places-to-find-free-datasets-for-data-science-projects-shared-article-from-dataquest/

# Evaluation

**Topic Description Points**

| Topic | Description | Points |
|---|---|---|
| **Interesting dataset** | Interesting dataset, loaded into your main file and used for analysis | **5 pts** |
| **main .ipynb script** | Well commented main .ipynb script | **6 pts** |
| **library .py script** | Well commented library .py script containing all of the relevant functions that you wrote for your analysis | **8 pts** |
| **key elements of program** | • Apply at least one *for* or *while* loop<br>• Apply at least one *if/elif/else* logical operators statements<br>• Sort the data in some way (biggest to smallest / by month / by year / by type / etc.)<br>• Perform at least three independent computations with the data (* / + - ** % //, etc.)<br>• Print at least 2 statements of output from your program to the terminal<br>• Create at least 3 plots showing your data and the analysis of your data that are well labeled and use different types of plotting styles<br>• Use the numpy, pandas, and matplotlib packages within your program | **8 pts per bullet**<br><br>**= 56 points total** |
| **.pdf** | The format and presentation of your final output of your program including analysis of your dataset and importance of your dataset | **25 pts** |

# Total = 100 pts (15% of your final grade)